ACEEE Int. J. on Network Security , Vol. 4, No. 1, July 2013

Data Accuracy Models under Spatio - Temporal Correlation with Adaptive Strategies in Wireless Sensor Networks

Jyotirmoy Karjee , H.S Jamadagni
Department of Electronic Systems Engineering , Indian Institute of Science, Bangalore, India
Email: {kjyotirmoy,hsjam}@cedt.iisc.ernet.in

***Abstract*—Wireless sensor nodes continuously observe and sense statistical data from the physical environment. But what degree of accurate data sensed by the sensor nodes collaboratively is a big issue for wireless sensor networks. Hence in this paper, we describe accuracy models of sensor networks for collecting accurate data from the physical environment under *two conditions. First condition:* we propose accuracy model which requires a priori knowledge of statistical data of the physical environment called Estimated Data Accuracy (EDA) model. Simulation results shows that EDA model can sense more accurate data from the physical environment than the other information accuracy models in the network. Moreover using EDA model, there exist an optimal set of sensor nodes which are adequate to perform approximately the same data accuracy level achieve by the network. Finally we simulate EDA model under the thread of malicious attacks in the network due to extreme physical environment. *Second condition:* we propose another accuracy model using Steepest Decent method called Adaptive Data Accuracy (ADA) model which doesn't require any a priori information of input signal statistics. We also show that using ADA model, there exist an optimal set of sensor nodes which measures accurate data and are sufficient to perform the same data accuracy level achieve by the network. Further in ADA model, we can reduce the amount of data transmission for these optimal set of sensor nodes using a model called Spatio-Temporal Data Prediction (STDP) model. STDP model captures the spatial and temporal correlation of sensing data to reduce the communication overhead under *data reduction strategies*. Finally using STDP model, we illustrate a mechanism to trace the malicious nodes in the network under extreme physical environment. Computer simulations illustrate the performance of EDA, ADA and STDP models respectively.**

***Index Terms*—Wireless senor networks, data accuracy, spatial correlation, adaptive filter**

I. Introduction

Recent progress in real time distributed system has made a drastic improvement for monitoring continuous data over wireless sensor networks. Such continuous monitoring of real data applications permits to observe in both time and space. In wireless sensor network, sensor nodes are deployed both in time and space to monitor the physical phenomenon of data (e.g temperature) from the physical environment [1]. For a specific time instant, sensor nodes collect the data in space domain and transmit it to the sink node. The major task of sensor nodes is to collect the data from the physical environment. Since the data collected by the sensor nodes are generally spatially correlated [2,3] among them, the sensor nodes need not require transmitting all the sensor readings to the sink node. Instead a subset of sensor reading is sufficient to transmit to the sink node maintaining a desired *accuracy* [7-11]. Thus exploring spatio-temporal correlated data to transmit a subset of sensor reading maintaining desired *estimated data accuracy* at the sink node is an emerging topic in wireless sensor network and is the key interest of this paper. This procedure can reduce a significant communication overhead and energy consumption in the network.

The first motivation of this paper is to develop accuracy models for the network to sense accurate data from the physical environment. To collect accurate data for the network, we develop accuracy models under two conditions. Firstly, we propose accuracy model called Estimated Data Accuracy (EDA) model which requires a priori knowledge of statistical data of the physical environment. We compare the performance of EDA model [7] with other information accuracy model [4-6] which illustrate that EDA model performs better than other models to select an optimal set of sensor nodes in the network. EDA model requires exact variances and covariances of the statistical data with prior information of the physical environment. But in practice, this type of prior information of signal statistics is difficult to get in real scenario and model it. Hence we propose another accuracy model called Adaptive Data Accuracy (ADA) model to overcome this difficulty. To the best understanding of authors, this is the first time, we propose ADA model which doesn't require any a priori information of statistical data of the environment and measures accurate data for the network. ADA model estimates a desired accuracy collaboratively using adaptive Steepest -Decent method [15] at the sink node from an optimal set of the sensor nodes instead of using all sensor nodes in the network. The data collected using ADA model is dynamic and doesn't require relying on historical information of data to estimate data accuracy at the sink node.

The second motivation of this paper is to reduce the communication overhead of optimal sensor nodes selected in the network using ADA model while maintain a certain degree of data accuracy. These optimal sensor nodes selected in the network using ADA model transmits a subset of sensor readings to the sink node to explore *data reduction strategies* [25]. In data reduction strategies, we use adaptive LMS filter to reduce the amount of data transmitted by each sensor nodes under spatially correlated data in the sensing

region. Under data reduction strategies, we propose a model called Spatio-Temporal Data Prediction (STDP) model which not only reduces the communication overhead but also have the learning and tracking capability to trace the internal variations of the statistical signal in the network. STDP model uses adaptive LMS filter both at the sensor nodes and the sink node. In STDP model, filter at the sink node does the *joint prediction* [25] to capture the spatial and temporal data correlation among the optimal sensor nodes in the sensing region.

Now it is crucial to explain the importance of STDP model. *Why STDP model is better than other data prediction models?* The answers of this question are illustrated as follows:

(i) In STDP model, sink node estimates a global weighted vector using LMS filter which captures the spatial and temporal data correlation among the sensor nodes in the wireless sensor region. Global weighted vector calculated at the sink node gives the information about the statistics of data in the network. Thus global weighted vector calculated at the sink node in STDP model estimates good data at the sink node for the network. But in other prediction models [23-25], sink node doesn't capture spatial and temporal data correlation features in the sensor region since a single weighted vector is considered individually for the respective sensor node.

(ii) If a node collects bad or malicious data, the estimated value of weighted vector calculated at the sink node degrades. In prediction model [25], weighted vector has to depend on a single node data collection. But STDP model is not restricted to predict the data from a single node like other prediction model. In STDP model, sink node has the knowledge of the statistics of whole data (spatio-temporal data correlation) of all sensor nodes using a global weighted vector. Thus STDP model does the *joint prediction* scheme at the sink node for data reduction which lags in literature [25].

(iii) In literature [25], filters at the sink node and the sensor nodes are always active. If the filter at the sensor node is always active, it consumes energy still it doesn't perform any transmission of data. But in STDP model, we have switching mode (like ON/OFF) mechanism to make the filters at sensor nodes and the sink node to be idle. Thus our STDP model can save more energy than the existing model.

(iv) In literature [26], a spatio-temporal model is illustrated where historical sensed data is taken to estimate sensor readings in current period. But in STDP model such historical data is not taken to estimate readings in current time period. In STDP model, spatio-temporal data is refreshed in a cyclic order after certain interval of time using a new global weighted vector calculated at the sink node.

The third and final motivation of this paper is to verify our propose models when the network is under the thread of malicious attacks. Since maximizing the network life time subjected to event constraint and information gathering to maximize the network life time subjected to energy constraints [18],[19] are discussed without verifying the data accuracy. Verifying data accuracy is essential before data aggregation as it degrades the accuracy level if some of the sensor nodes gets malicious [7],[12] due to extreme physical environment like heavy rain fall etc. Thus inaccurate data aggregated with the other correct data results poor data aggregation and reduces data accuracy level at the sink node. We perform EDA model under the tread of malicious attacks. We simulate and compare EDA model when the network is under thread of attacks as well as when the network is good (not under thread of attacks). EDA model estimates data accuracy under the thread of malicious nodes but don't incorporate to trace the number of malicious nodes in the network. Therefore, finally in this paper, we propose a mechanism to find the number of malicious nodes in the network if any using STDP model.

The rest of the paper is given as follows. In section II, we explore the motivation and problem definitions of our work. In section III, we explain briefly the accuracy model which requires a priori knowledge of signal statistics of the environment. Further in section IV, we explain accuracy model which doesn't require a priori knowledge of signal statistics and the data reduction model. In section V, we perform the simulation as well as validation and finally conclude our work in Section VI.

II. Overview Of Approach and Problem Definations

The purpose and motivation of this paper is explained in threefold which are as follows:

(i) In wireless sensor networks, sensor nodes sense statistical data from the physical environment and transmit it to the sink node. But how much accurate data is collected by the sensor nodes collaboratively is a big issue with respect to quality of services. Hence in this paper, we develop data accuracy models which can extract accurate data from the physical environment. Again these data accuracy models are categorized under two situations. In one situation, accuracy model called Estimated Data Accuracy (EDA) model is proposed where a priori knowledge of statistical data is known. EDA model performs better than other information accuracy models [4-6] and can still be meeting by an optimal set of sensor nodes rather than taking all the sensor nodes in the network maintaining a desired data accuracy. In EDA model, variance or covariance of the sensed data is assumed to be known. In other words, we have a priori knowledge of input signal statistics of the environment in EDA model. In another situation, we propose Adaptive Data Accuracy (ADA) model which doesn't require any a priori knowledge of input signal statistics of the environment. ADA model has the capability for tracing continuous data stream for a regular time interval to estimate the required signal. ADA model also extract accurate data from the physical environment and satisfies the criteria to find an optimal set of sensor nodes maintaining the desired data accuracy.

(ii) One of the major tasks of sensor nodes in wireless sensor network is to transmit a subset of sensor readings to the sink node estimating a desired data accuracy. To fulfill this task, at first an optimal set of sensor nodes are selected in the network using ADA model maintaining a certain

ACEEE Int. J. on Network Security , Vol. 4, No. 1, July 2013

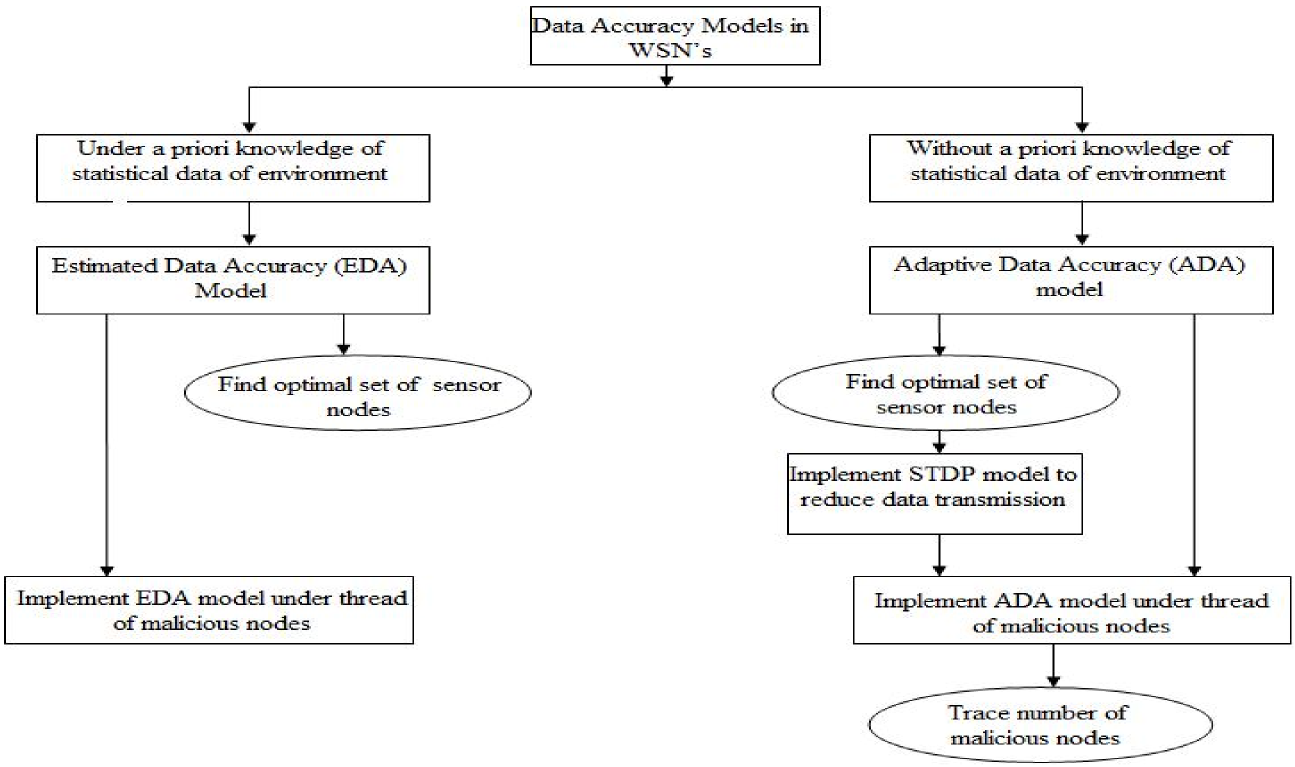

Figure1. System Architecture for Data Accuracy Models in Wireless Sensor Networks

degree of data accuracy. Then these optimal set of sensor nodes transmits a subset of sensor readings to the sink node. This reduces the amount of data transmission and communication overhead. This can be done using methodology called Spatio-Temporal Data Prediction (STDP) model. STDP model is developed to capture the signal statistics when the data are correlated among the sensor nodes in the wireless sensor network. This STDP model performs a *joint prediction* scheme which can learn and track the internal variation of the signal statistics to adopt itself with the environment. STDP model reduces the communication overhead in the wireless sensor network based on data reduction strategies.

(iii) Finally, we verify our EDA and ADA models under the threads of malicious attacks. We evaluate the performance of EDA model by introducing some malicious nodes in the network and compare with good network. In EDA model, we don't have any methodology to find the number of malicious nodes in the network. But in ADA model, we are able to find a mechanism using STDP model which not only have the capability to trace the number of malicious nodes in the network but also have the learning as well as tracking capability. We summarize our motivation of work in this paper given in Fig.1. In the later sections, we discuss data accuracy models with and without a priori knowledge of signal statistics of data respectively.

III. Data Accuracy Model with a Priori Knowlegge of Signal Statistics

In this section, we develop a mathematical foundation of data accuracy model called Estimated Data Accuracy (EDA) model which requires a priori knowledge of statistical data of the physical environment. Moreover we perform EDA model under the thread of malicious attacks.

A. Estimated Data Accuracy Model

Initially, we deploy randomly M sensor nodes in a sensing region. Assuming M sensor nodes collaboratively senses the physical phenomenon of desired signal d. For simplicity we call these M sensor nodes as Clients .We construct a mathematical model to estimate the observed data at the sink node. Sink node is like as server which is responsible for collecting the observation made by M sensor nodes to estimate $\hat{d}$ from d . The error signals [13, 14] can be defined as

$$\tilde{d} = (d - \hat{d}) \tag{1}$$

We find $\hat{d}$ by minimizing the mean square error from the expectation of signal $\tilde{d}^2$ as follows.

$$\min_{\hat{d}} E(\tilde{d})^2 \tag{2}$$

The observation done by each sensor node i is given as

$$u_i = d_i + v_i \text{ where } i \in M \tag{3}$$

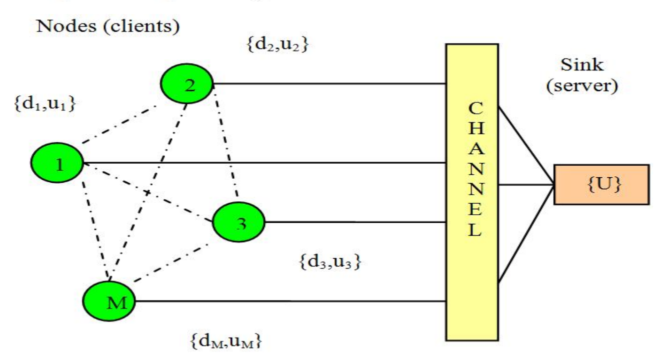

Figure 2: Architecture of System Model

DOI: 01.IJNS.4.1.1193

Assuming uncoded transmission for the observed signal sensed by sensor nodes in the wireless sensor network. Each sensor node i sends a scaled version Y_i [4] of the observed signal u_i to the sink node with power constraint P in the network. Hence transmitted signal in the network is given as

$$Y_i = \sqrt{\frac{P}{\sigma_{d_i}^2 + \sigma_{v_i}^2}} u_i \quad \text{where} \quad \psi_i = \sqrt{\frac{P}{\sigma_{d_i}^2 + \sigma_{v_i}^2}} \quad \text{for} \quad i \in M$$

$$\text{Hence} \quad Y_i = \psi_i u_i \qquad \text{for} \quad i \in M$$

The signal Y_i transmitted by each sensor node i is sent to the sink node through additive white Gaussian noise (AWGN) channel [6],[13] in the network. Sink node store the received signal in U matrix as shown in Fig. 2 for all sensor nodes in the network as

$$U = \psi X \tag{4}$$

$$\text{for}\, U = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ Y_M \end{pmatrix}, \psi = \begin{pmatrix} \psi_1 & 0 & . & 0 \\ 0 & \psi_2 & . & . \\ . & . & . & . \\ 0 & . & . & \psi_M \end{pmatrix} \text{and} \quad X = \begin{pmatrix} u_1 \\ u_2 \\ . \\ u_M \end{pmatrix}$$

$$\text{where} \quad i = 1, 2 \ldots\ldots M$$

Sink node retrieve the signal to the estimate $\hat{d}$ of d . We assign $\hat{d}$ as a random variable of function U to recover the estimate $\hat{d}$ of observed signal made by M sensor nodes collaboratively at the sink node.

$$\hat{d} = h(U) \tag{5}$$

Thus the mean square error is represented as

$$\min_{h(U)} E(\tilde{d})^2 \tag{6}$$

We take $h(U)$ for the subclass of affine functions [15] of U as

$$\hat{d} = h(U) = (WU + g)$$

where W is matrix and g is a scalar quantity. Hence affine estimator of d is taken unbiased with $E(\hat{d})$ = 0 and $E(\hat{d}) = WE(U) + g = g$ .For a linear estimator, we have g = 0 to get

$$\hat{d} = WU \tag{7}$$

We find the optimal value of W at the sink for $\hat{d}$ to minimize

$$\min_{W} E(d - WU)^2 \tag{8}$$

We find the optimal value of W for the estimate of $\hat{d}$ using orthogonal [15] function. The vector U is orthogonal to the error signal $(\tilde{d})$ .To get the optimal value of W for the estimator at the sink node, we define a linear model [7] for (4) as

$$U = \psi(Zd + V) \tag{9}$$

$$U = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ Y_M \end{pmatrix} = \begin{pmatrix} \psi_1 & 0 & . & 0 \\ 0 & \psi_2 & . & . \\ . & . & . & . \\ 0 & . & . & \psi_M \end{pmatrix} \left[\begin{pmatrix} Z_1 \\ Z_2 \\ . \\ Z_M \end{pmatrix} d + \begin{pmatrix} v_1 \\ v_2 \\ . \\ v_M \end{pmatrix} \right]$$

for zero mean random vector $\{d, U\}$ for some matrix Z . V is a zero mean random noise vector with known covariance matrix $E(VV^T) \approx \sigma_V^2 I$. The covariance matrix of d is also known $E(dd^T) \approx \sigma_d^2 I$ and $\{d, V\}$ are uncorrelated. The linear least mean square estimator works according to orthogonal function as $E\left(U^T \tilde{d}\right) = E\left(U^T\left(d - WU\right)\right) = 0$. Using (9) we write W as

$$W = \frac{E(U^T d)}{E\left(U^T U\right)} = \frac{\sigma_d^2 Z^T \psi^{-1}}{\left(\sigma_d^2 Z^T Z + \sigma_V^2\right)} \tag{10}$$

Using (10) in (7), we get the linear least mean square estimator of d given U is

$$\hat{d} = \frac{\sigma_d^2 Z^T \psi^{-1}}{\left(\sigma_d^2 Z^T Z + \sigma_V^2\right)} U = \frac{Z^T}{\left(Z^T Z + \sigma_V^2 / \sigma_d^2\right)} X \tag{11}$$

Hence we use linear least mean square estimator of d given X for M sensor nodes in the network is given as

$$\hat{d}(M) = \frac{1}{\left(M + \sigma_V^2/\sigma_d^2\right)} \sum_{i=1}^{M} X_i = \frac{1}{\eta} \sum_{i=1}^{M} X_i \text{ ,where } \eta = \left(M + \frac{\sigma_V^2}{\sigma_d^2}\right) \tag{12}$$

We define mean square error between d and $\hat{d}(M)$ to find the estimated data accuracy [7] for M sensor nodes in network as

$$D(M) = E[d - \hat{d}(M))]^2 = E[d^2] - 2E[d\hat{d}(M)] + E[\hat{d}(M)^2] \tag{13}$$

The normalized [6,10] data accuracy $D_A(M)$ for M sensor nodes in the network is given as

$$D_A(M) = 1 - \frac{D(M)}{E[d^2]} = \frac{1}{E[d^2]}[2E[d\hat{d}(M)] - E[\hat{d}(M)^2]] \tag{14}$$

The normalized data accuracy $D_A(M)$ for M sensor nodes in the sensor region can be implemented in spatial correlation model [7] in the network.

Now we model a spatially correlated physical phenomenon of sensed data for M sensor nodes as a joint Gaussian random variable (JGRV's) [4,5] as : $E[d_i] = 0$ and $\text{var}[d_i] = \sigma_d^2$ .The covariance between d and d_i is $\text{cov}[dd_i] = E[dd_i] = \sigma_d^2 K(dis_{ij})$.

Similarly the covariance between d_i and d_j is $\text{cov}[d_i d_j] = E[d_i d_j] = \sigma_d^2 K(dis_{ij})$. The covariance model [16] $K(dis_{i,j})$ where $dis_{i,j} = \| d_i - d_j \|$ represents the Euclidian distance between node i and j .The covariance

DOI: 01.IJNS.4.1.1193

function is non-negative and decrease monotonically with the Euclidian distance $dis_{i,j} = \| d_i - d_j \|$ with limiting values of 1 at $dis = 0$ and of 0 at $dis = \infty$. We take the power exponential model [17] i.e $K(dis_{i,j}) = e^{-(dis_{(i,j)}/\theta)}$ for $\theta > 0$ where θ is called as 'Range parameter'. 'Range parameter' controls the relation between the distance among sensor nodes (i, j) and the correlation coefficient $\rho(i, j)$. Using the correlation model, we get $\rho_{d_i,d} = e^{-(dis_{(d,i)}/\theta)}$ and $\rho_{d_i,d_j} = e^{-(dis_{(i,j)}/\theta)}$. Using (3) and (12) in (14), we get the EDA model in the network as

$$D_A(M) = \frac{1}{\eta}\left(2\sum_{i=1}^{M} e^{-(dis_{(d,i)}/\theta)}\right) - \frac{1}{\eta^2}\left(\sum_{i=1}^{M}\sum_{j\neq i}^{M} e^{-(dis_{(i,j)}/\theta)}\right) - \frac{1}{\eta^2}\left(\frac{M\sigma_d^2 + \sum_{i=1}^{M}\sigma_{V_i}^2}{\sigma_d^2}\right) \quad (15)$$

B. EDA Model under the tread of malicious attacks

Data gathering or data aggregation are the traditional procedure subjected to energy constraints [19] and maximizing [18] network life time. These procedures are done without verifying the data accuracy before data aggregation in the network. Hence without verifying the data accuracy before data aggregation cause problem if some of the senor nodes get malicious [7] in the network. The sensor node gets malicious due to extreme physical environment e.g heavy rainfall or snow fall etc. Malicious nodes sense inaccurate data readings and transmit the inaccurate data to the sink node. Sink node aggregate inaccurate data with the other correct data send by the sensor nodes. Thus sink node estimate inaccurate data reading which results poor data gathering in the network. We perform EDA model under such situation when the network is under the thread of malicious attack.

IV. Data Accuracy Model without a Priori Knowlegge of Signal Statistics

In this section, at first we construct the mathematical foundation to select the optimal sensor nodes using a model called Adaptive Data Accuracy (ADA) model. This ADA model doesn't require a priori knowledge of input signal statistics to estimate the required signal. These optimal sensor nodes are selected from the network using ADA model to perform data transmission maintaining desired data accuracy. Data transmission can be further reduced using a methodology called Spatio-Temporal Data Prediction (STDP) model. STDP also captures the spatial and temporal correlation of data among the sensor nodes to reduce the communication overhead in the network using LMS (adaptive) filter. Finally using STDP model, we find a mechanism to trace the malicious nodes in the network.

We consider M sensor nodes randomly distributed over a wireless sensor network. When a query is requested from the sink node to the sensor nodes, the sensor nodes start sensing the physical phenomenon of data e.g temperature from the environment and transmit data to the sink node. For the simplicity of system model, we called these sensor nodes as clients and the sink node as the central server as shown in Fig. 2. Hence clients and server performs the following roles in wireless sensor networks for data transmission.

Clients: Each sensor node i can sense and observe the physical phenomenon of data in the wireless sensor network. The observation made by each sensor node to collect the continuous block [20] of data samples up to samples over a window frame of time interval is given as

$$u_i = \{u_i^1, u_i^2 \ldots . u_i^N\} \ (1 \times N) \quad \text{where } i \in M \qquad (16)$$

The corresponding scalar measurement (desired signal) done by each sensor node is given as

$$d_i = u_i w_0 + v_i \quad \text{where } i \in M \qquad (17)$$

where w_0 is $(N \times 1)$ an initial weighted vector with unknown matrix in the client side. v_i is the temporal and spatial uncorrelated white noise. Each sensor node i transmits u_i observation to the sink node through additive white Gaussian noise (AWGN) channel [6],[13] in the wireless sensor network.

Server: Sink node restores the observed data received from all the sensor nodes in U matrix and the corresponding desired signal in d matrix in the network as follows

$$U = col\{u_1, u_2 \ldots\ldots\ldots u_M\} \ (M \times N) \qquad (18)$$

$$d = col\{d_1, d_2 \ldots\ldots\ldots\ldots d_M\} \ (M \times 1) \qquad (19)$$

A. Adaptive Data Accuracy Model

We propose a model called Adaptive Data Accuracy (ADA) model which doesn't require a priori knowledge of signal statistics and can trace the continues data stream for a regular interval of time under spatially correlated data in the sensor region .

Since M sensor nodes are randomly deployed, the observed data sensed by the sensor nodes are spatially correlated among them in the sensor network. Hence we can reduce the number of sensor nodes while maintaining approximately the same data accuracy level which we achieve by M sensor nodes, since the observed data are spatially correlated among them. We perform Minimum Mean Square Estimation (MMSE) [14] with adaptive [21]approach at the sink node to reduce the number of sensor nodes subjected to data accuracy for the spatially correlated observed data sensed by M sensor nodes in the network.

Since our aim is to minimize the cost function [22] as

$$J(w) \underset{\min w}{=} E\| d - U\mathbf{w} \|^2 \qquad (20)$$

to get an optimal solution $\mathbf{w}_0$ using normal equation $\Re_{du} = \Re_{uu}\ \mathbf{w}_0$. Here U is a $(1 \times M)$ row observation vector, $\mathbf{w}$ is another weighted vector of $(M \times 1)$ matrix and d is a scalar desired signal calculated at sink node.

Expanding (20), we get the Minimum mean square estimation (MMSE) for the ADA model given as

$$J(w) = \sigma_d^2 - \Re_{du}^T \mathbf{w} - \mathbf{w}^T \Re_{ud} + \mathbf{w}^T \Re_{uu} \mathbf{w} \quad (21)$$

where $\Re_{uu} = E[U^T U]$; $\Re_{du} = E[U^T d]$; $\sigma_d^2 = E|d|^2$

Using adaptive Steepest Decent method [15],[22],sink node find a global statistical information $\{\Re_{uu}, \Re_{du}\}$ for the spatially correlated data in the network as follows.

$$\Re_{uu} = E[U^T U] = \begin{pmatrix} E[u_1 u_1] & \dots & E[u_1 u_M] \\ \vdots & \ddots & \vdots \\ E[u_M u_1] & \cdots & E[u_M u_M] \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{u_1}\sigma_{u_1}\rho_{u_1u_1} & \dots & \sigma_{u_1}\sigma_{u_M}\rho_{u_1u_M} \\ \vdots & \ddots & \vdots \\ \sigma_{u_M}\sigma_{u_1}\rho_{u_Mu_1} & \cdots & \sigma_{u_M}\sigma_{u_M}\rho_{u_Mu_M} \end{pmatrix}_{(M\times M)} \quad (22)$$

Similarly

$$\Re_{du} = E[U^T d] = \begin{pmatrix} E[du_1] \\ E[du_2] \\ .. \\ .. \\ E[du_M \end{pmatrix} = \sigma_d \begin{pmatrix} \sigma_{u_1}\rho_{du_1} \\ \sigma_{u_2}\rho_{du_2} \\ . \\ . \\ \sigma_{u_M}\rho_{du_M} \end{pmatrix}_{(M\times 1)} \quad (23)$$

We model spatially correlated data as a Joint Gaussian Random Variable (JGRV)'s [4],[5] as follows : $d = \frac{1}{M}\sum_{i=1}^{M} d_i$, $E|d|^2 = \sigma_d^2$ and $E[u_i] = \frac{1}{N}\sum_{j=1}^{N} u_{ij}$ where $i=1,2,3\ldots M$ nodes and $j=1,2,3,\ldots N$ samples in each node ; standard deviation of $u_i = \sigma_{u_i}$, standard derivation of $d = \sigma_d$ for $i = 1, 2, \ldots M$ nodes. The covariance between u_i and u_j is $Cov[u_i u_j] = E[u_i u_j] = \sigma_{u_i}\sigma_{u_j}\rho_{u_iu_j}$ where $\rho_{u_iu_j}$ is the correlation coefficient between d and u_i for $i = 1,2,\ldots\ldots M$. We define correlation model [4] $K(.)$ as $K(dis_{i,j}) = \rho_{u_iu_j}$ where $dis_{i,j} = \| u_i - u_j \|$ is the Euclidian distance between the sensor nodes i and j .We adopt power exponential model [16, 17] in correlation model as $K^{PE}(dis_{i,j}) = e^{-(dis_{i,j}/\theta)}$ for $\theta > 0$ where θ is a 'Range Parameter' [4],[7].Thus we get $\rho_{u_iu_j} = e^{-(dis_{i,j}/\theta)}$ and $\rho_{du_i} = e^{-(dis_{d,i}/\theta)}$.

We start from an initial guess for $\mathbf{w}$ and derive a procedure in recursive manner until it converges to $\mathbf{w}_0$. Hence the weighted vector for ADA model is given as

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mu[\Re_{du} - \Re_{uu}\mathbf{w}_{k-1}] \ (Mx1) \quad (24)$$

Where $\mathbf{w}_k$ is the weighted vector with new guess having k iterations. $\mathbf{w}_{k-1}$ is a old guess for $(k-1)$ iterations and $\mu > 0$ is a positive step size parameter [15],[22],[27] which is calculated under spatial correlation of data among sensor nodes in the network as follows

$$\Re_{uu} = V \Lambda V^T \quad (25)$$

where Λ is the diagonal positive entries of eigen values and V is a unitary matrix satisfying $VV^T = V^T V = I$. We pick the largest eigen value $(\lambda_{\max})$ to find μ according to [15] as given as

$$0 < \mu \le (2/\lambda_{\max}) \quad (26)$$

Putting optimal weighted vector calculated from (24) in cost function $J(w)$ equation (21), we get the normalized data accuracy at the sink node for the network is given as

$$Accuracy(M) = 1 - \frac{J(w)}{\sigma_d^2} \quad (27)$$

This normalized $Accuracy(M)$ using ADA model calculated at the sink node is used for finding the optimal number of sensor nodes in the network subjected to data accuracy. Thus ADA model doesn't require a priori knowledge of input signal statistics to calculate the data accuracy of signal at the sink node and can trace the continuous data stream for a regular interval of time in the sensor network.

B. Spatio-Temporal Data Prediction Model

Since we select optimal sensor nodes using ADA model in the network maintaining a certain level of data accuracy, these optimal sensor can perform data transmission in the network. We develop a methodology called Spatio-Temporal Data Prediction (STDP) model by which we can further reduce the data transmission for these optimal sensor nodes in the network. STDP captures the spatial and temporal correlation of data to reduce the communication overhead among the sensor nodes based on *data reduction strategies*. In data reduction strategies, sensor nodes only transmit a subset of data stream to the sink node instead of transmitting the whole data stream. This reduces communication overhead in the network. STDP model performs a *joint prediction* scheme to capture the spatial correlation among sensor nodes to reduce the communication overhead. Moreover our approach doesn't require any a priori knowledge of input signal statistics and have the learning as well as tracking capability to trace the internal variation of the signal statistics.

When a query is requested from the sink node (server) to all the sensor nodes (clients), STDP model starts transmitting the data to the sink node for *joint prediction* of data in the network as follows:

***Phase I** Client:* Each sensor node i in the client side transmits the spatially correlated data u_i according to (16) and the corresponding measured data d_i according to (17) (along with initial weighted vector w_0^{int}) to the sink node. w_0^{int} is an unknown vector $w_0^{\text{int}} = col\{1,1,\ldots\ldots\ldots,1\}/\sqrt{N}$ [22]. Initially at this moment, the filter at each sensor node and the filter to be used at the sink node are kept ideal in the network.

***Phase II** Server:* Sink node store the received u_i observation transmitted from i sensor nodes in U matrix and

DOI: 01.IJNS.4.1.1193

d_i in d matrix according to (18) and (19) respectively. Using adaptive Steepest-decent method [15],[22] , sink node find another global statistical information {R_{UU}, R_{DU}} for the spatially correlated data in the network as follows

$$R_{UU} = E[U^T U]\ (N\times N) \text{ and } R_{DU} = [U^T d]\ (N\times 1) \quad (28)$$

Hence using this method [22], the estimated global weighted vector calculated at the sink node is given as

$$w_k^{Glob} = w_{k-1}^{Glob} + \mu\sum_{i=1}^{M}\left(R_{DU,k} - R_{UU} w_{k-1}\right)\ (N\times 1) \quad (29)$$

According to instantaneous approximations [22], the global statistical information can be written as

$$R_{UU,i} = u_i^T u_i\ (N\times N) \text{ and } R_{DU,i} = u_i^T d_i\ (N\times 1) \quad \text{for } i\in M \quad (30)$$

Hence using instantaneous approximations, the global weighted vector w_k^{Glob} (29) can be modified as

$$w_k^{Glob(IA)} = w_{k-1}^{Glob} + \mu\sum_{i=1}^{M}\left(u_i^T d_i - \left(u_i^T u_i\right) w_{k-1}^{Glob}\right) \quad (N\times 1) \quad \text{for } i\in M \quad (31)$$

Similarly using LMS filter [22], the global weighted vector w_k^{Glob} calculated at the sink node is given as

$$w_k^{Glob(LMS)} = w_{k-1}^{Glob} + \mu\sum_{i=1}^{M} u_i^T\left(d_i - u_i w_{k-1}^{Glob}\right) \quad (N\times 1) \quad \text{for } i\in M \quad (32)$$

Now $w_k^{Glob(LMS)}$ is used to calculate the prediction filters (for M nodes) at sink as $y_{Sink} = U w_k^{Glob(LMS)}$. This realizes that at this moment, the prediction filters at sink node are active to calculate y_{Sink} and prediction filter (y_i) used at each sensor node i is still kept idle. Finally we use y_{Sink} to calculate the prediction error at sink node as

$$error^{Glob} = [d - y_{Sink}] \quad (33)$$

We define a user defined error threshold (α) value to satisfy these two conditions defined as follows:

- If the $error^{Glob}$ (scalar value) calculated at the sink node for the respective node is greater than the error threshold (α) value, then the corresponding sensor node still continue to send the data to the sink node. This makes the filter at the sink node to adopt well for the received data transmitted from the corresponding node and goes to adaptive mode. In this situation, filter at the sink node (server) for this node is active and the filter for this corresponding node (client) is still kept ideal.
- But if the $error^{Glob}$ for the respective node is less than the error threshold (α) value, the global weighted vector $w_k^{Glob(LMS)}$ calculated at the sink node is transferred to the corresponding sensor node in the network. This transmission of $w_k^{Glob(LMS)}$ is like a *request* from the sink node to sensor node for stopping the transmission of data [23]. Once $w_k^{Glob(LMS)}$ is transferred from the sink node to sensor node, filter residing at sink node for the sensor node goes to ideal and it goes to prediction mode , finally filter residing for the corresponding sensor node is yet to become active.

***Phase III** Client:* Once $w_k^{Glob(LMS)}$ is received by each sensor node (conditioned *error* calculated at the sink node for the respective node is less than the threshold), it (client) uses $W_k^{Glob(LMS)}$ to calculate its new weighted vector, filter and error. Since u_i observation is sensed by sensor node i , the desired signal scalar value calculated by each sensor node is given as

$$d_i^{new} = u_i w_k^{Glob(LMS)} + v_i \quad \text{where } i\in M \quad (34)$$

Hence the new updated weighted vector calculated at each sensor node i in the network is given as

$$w_{i,k}^{new} = w_{i,k-1} + \mu\left(u_i^T\left(d_i^{new} - u_i w_{i,k-1}\right)\right) \quad (N\times 1) \quad \text{for } i\in M \quad (35)$$

Now each sensor node finds its individual filter update value as $y_i^{new} = u_i w_{i,k}^{new}$ and finally the scalar error value calculated for each sensor node i in the network is given as

$$error_i^{new} = \left[d_i^{new} - y_i^{new}\right] \quad (36)$$

Again using another user defined error threshold (β) value, we illustrate two conditions:

- If $error_i^{new}$ calculate at the sensor node is greater than the error threshold value (β) , observation u_i sensed by it is still transmitted to the sink node. At this time filter at node is active and goes to adaptive mode. At this stage, filter at the sink for the corresponding node is set idle.
- If $error_i^{new}$ calculate at the sensor node is less than the error threshold value (β) , the sensor node stop transmitting the observation u_i to the sink node. Once data transmission is stopped, sensor node i transfer it's $w_{i,k}^{new}$ to the sink node. This transmission of $w_{i,k}^{new}$ is like a *response* from the sensor node to the sink node that transmission of data is stopped. At this moment filter at the node is set ideal and goes to prediction mode. Sink node utilizes $w_{i,k}^{new}$ transmitted from each sensor node to

DOI: 01.IJNS.4.1.1193

track the signal statistics of each sensor node i in the network. Once $error_i^{new} \geq \beta$, Repeat the same process as we explained in ***Phase I.***

Thus using these three phases of STDP model, we perform the *data reduction strategies* to reduce the communication overhead for the optimal sensor nodes in the network. STDP model can track and learn the internal variation of the signal statistics without requiring a priori knowledge of the environment .Moreover it does the *joint prediction* scheme for the optimal sensor nodes to capture spatial and temporal correlation of data among sensor nodes in the network.

C. Tracing Malicious Nodes in the Network

In EDA model, we perform data accuracy for the network when some of the sensor nodes get malicious. But does not incorporate to trace and discard the malicious nodes from the network to get better data accuracy. Hence using STDP model, we find a novel methodology to trace the malicious nodes in the network. Since M sensor nodes are randomly deployed over a sensor region, we assume some of the nodes get malicious due to extreme physical environment. But we don't know the extract number of malicious nodes out of M sensor nodes in the network. Our motivation is to trace these malicious nodes in the network. The node is malicious or not depends upon the statistical behavior of observation u_i . We repeat the same procedure explained in STDP model of *Phases-I-III* where weighted vector $w_{i,k}^{new}$ gives the statistical information of each node to trace the malicious behavior. If the node is malicious, then the statistical value $(w_{mal\_i,k}^{new})$ of that node is different from the normal [7] nodes .Finally we can trace the malicious nodes and discard it from the network to get better data accuracy and data aggregation under spatially correlated data.

V. Performance Evaluations and Validations

In this section, the simulation results are performed using matlab to validate the effectiveness of our proposed data accuracy models under a priori and without a priori knowledge of signal statistics of the physical environment respectively. To perform the simulations, a sensing region of $4m \times 4m$ grid based wireless sensor topology is taken with a sink node in the network. We deploy ten sensor nodes in the sensing region. Each sensor node can sense the observations (e.g temperature) from the physical environment and transmit it to the sink node.

A. Performance Evaluation of Data Accuracy Model with a Priori Knowledge of Signal Statistics.

Here we simulate Estimated Data Accuracy (EDA) model which require a priori knowledge of information about the physical environment and compare with the other information accuracy models. Moreover we also simulate EDA model under the thread of malicious attacks.

Performance of EDA model compare to other accuracy models:: We perform the simulation for EDA model with ten sensor nodes as shown in Fig. 3. Result shows that EDA can sense more accurate data and performs better than other information accuracy models [4-6]. Moreover as we keep on increasing the number of sensor nodes to ten, the data accuracy remains approximately constant in the sensing field. It shows that six sensor nodes are sufficient for achieving the same estimated data accuracy level instead of deploying ten sensor nodes. Hence it is unnecessary to choose all the ten sensor nodes as six sensor nodes are sufficient to perform the communication process maintaining the desired data accuracy making rest of the sensor nodes to be in sleep mode. Thus an optimal set of sensor nodes are sufficient to maintain the desired data accuracy level using EDA model in the network.

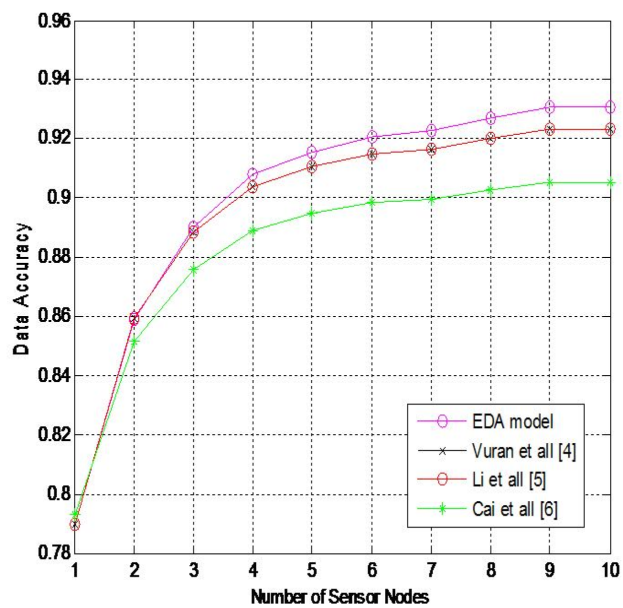

Figure 3. Number of sensor nodes versus data accuracy

Performance of EDA model under the thread of malicious attacks: We assume some of the senor nodes become malicious due to extreme physical environment like heavy rainfall etc. In such situation noise variances of malicious nodes are much higher than normal nodes. Normal nodes are good nodes and not under the thread of malicious attack. Thus in this simulation set up, we compare two deployment strategy for EDA model as shown in Fig 4. In the first strategy, initially we deploy five sensor nodes and keep adding sensor nodes to ten sensor nodes. These nodes are normal nodes. In another node deployment strategy, initially we deploy five sensor nodes in similar way but out of five sensor nodes, assuming two sensor nodes are malicious. We keep going on adding sensor nodes till we get ten sensor nodes in the network. We compare these two deployment strategies of EDA model and conclude that sink node estimates more accurate data when there are normal nodes in the network. But if the sensor nodes are under the thread of malicious nodes, the sink node estimates inaccurate data and performs poor data gathering for the network.

DOI: 01.IJNS.4.1.1193

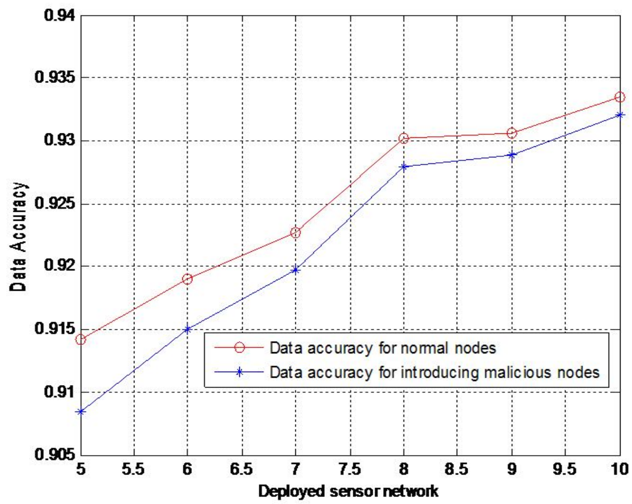

Figure 4. Comparison for data accuracy under normal nodes and under the thread of malicious nodes in EDA model

B. Performance Evaluation of Data Accuracy Model without a Priori Knowledge of Signal Statistics

We simulate for an accuracy model called adaptive Data Accuracy (EDA) model which doesn't require a priori knowledge of information about the physical environment. Using ADA model, we show that there exist an optimal set of sensor nodes which are sufficient for achieving the desired data accuracy level. Thus optimal sensor nodes are selected using ADA model in the network to perform the communication process maintaining a certain degree of data accuracy. Further Spatio-temporal Data Prediction (STDP) model is used to reduce the data transmission for these optimal set of sensor nodes. Finally using STDP model we find a mechanism to trace the number of malicious nodes in the network.

Since ADA and STDP models have the ability to trace the internal variations of the signal statistics to adopt itself with the physical environment, we consider such a tropical environment for our experiment where the variations of signal statistics (e.g temperature) are much more for certain duration. ADA and STDP models can work well to trace the variations of signal statistics for tropical desert area like Jaisalmer (Rajasthan-India). On 26th January 2012, the minimum and maximum temperatures recorded in jaisalmer are 7^0 Celsius and 22^0 Celsius respectively according to [28]. Such variations of temperature for a particular duration are the subject of interest to measure the variation of signal statistics of temperature rather than measuring the temperature variation of room temperature using ADA and STDP models. Since the temperature variation of room temperature is very less say for example 26^0 Celsius to 30^0 Celsius for a particular duration. Hence for our simulation purposes, we generate random data (temperature) using matlab which is sensed by sensor nodes to validate our results for ADA model and STDP model respectively. Thus the sensor nodes reported random (temperature) data once every 15 minutes recorded over one day (26th Jan 2012) on jaisalmer. Another example is to choose subtropical highland climate like Mawsynram[1] and cherrapungi ,(Meghalaya- India) where our model can works well to trace the variations of the signal statistics (measure the rainfall in mm) in the sensing region.

Performance of ADA model to select optimal sensor nodes: In this simulation, we estimate the data accuracy of the signal statistics at the sink node for all the deployed ten sensor nodes in the sensing region using ADA model. In Fig 5, we perform data accuracy of the signal statistics at the sink node for the ADA model with respect to number of sensor nodes in the network. Simulation results shows that about six sensor nodes are sufficient to perform approximately the same data accuracy as achieve by the ten sensor nodes in the sensing region. Thus an optimal (six) sensor nodes can perform data accuracy using ADA model instead of using ten sensor nodes in the network. We choose about six sensor nodes which are almost close to the sink node are eligible to perform the data transmission in the network maintaining a desired accuracy level using ADA model. Thus reducing the number of sensor nodes or selecting optimal sensor nodes for data transmission maintaining a desired accuracy in the network can reduce the communication overhead and increase the lifetime of the network.

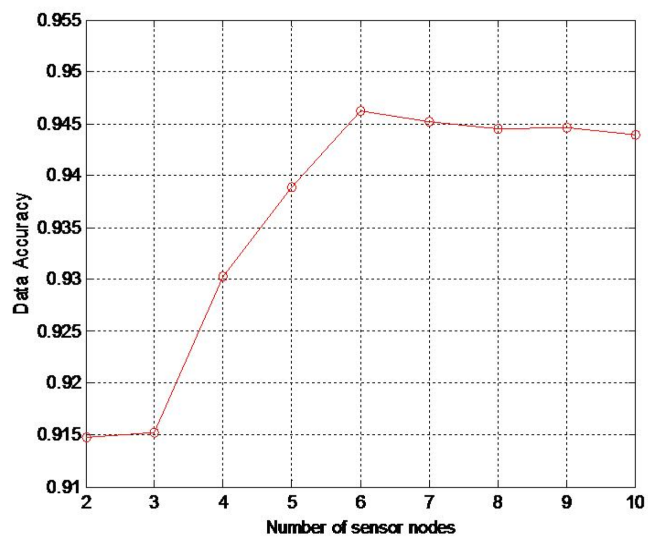

Figure 5. ADA model: data accuracy vs number of nodes

Performance of STDP model to reduce communication overhead: From the previous simulation setup, we select optimal (six) sensor nodes instead of taking ten sensor nodes for performing the communication process in the network using ADA model maintaining a desired data accuracy. Further we can reduce the data transmission using STDP model for these optimal set of sensor nodes. Hence in this simulation setup, we illustrate the performance analysis of STDP model.

Since using ADA model, we select about six sensor nodes (optimal sensor nodes selected) to perform the data transmission instead of using ten sensor nodes, we assume these optimal sensor nodes are with node Id's 2,4,5,7,9 and 10 selected for data transmission of signal statistics. These sensor nodes are chosen such that they are close to the sink node in the sensing region. Further using STDP, we can reduce the communication overheadof the data transmission

for these six sensor nodes to the sink node in the network. To analysis our simulation results in Fig.6 using STDP model, we report data transmission percentage of sensor readings for optimal number of nodes selected with respect to a varying error threshold value (β) . For each sensor node, a prediction error is calculated for each data stream of (N block) sensor readings to transmit. If the prediction error is greater than the error threshold value, the respective data stream of (block) sensor readings are transmitted by each sensor node to the sink node. Instead of transmitting all the data streams, a subset of data stream sensor readings for each sensor node is delivered to the sink node using data reduction strategies. Moreover Fig. 6 also shows statistical variations of signal for sensor nodes in the network. The statistical data streams among sensor nodes are almost similar because the data streams are spatially and temporally correlated among them. Thus subsets of data streams are transmitted by these optimal sensor nodes using STDP model to further reduce the data transmission in the network.

In Fig 7, we compare the percentage of data transmission for data stream block size $N = 4$ and data stream block size $N = 5$ for sensor node Id-2 with respect to error threshold value (β) . The simulation result shows that if we transmit data stream block size of, we can reduce the percentage of transmission cost effectively than transmitting data stream block size of. Another conclusion is drawn as the data stream block size is small, tracking and learning of statistical signal is easier whereas if we use data stream block size larger, a better estimation of statistical is performed.

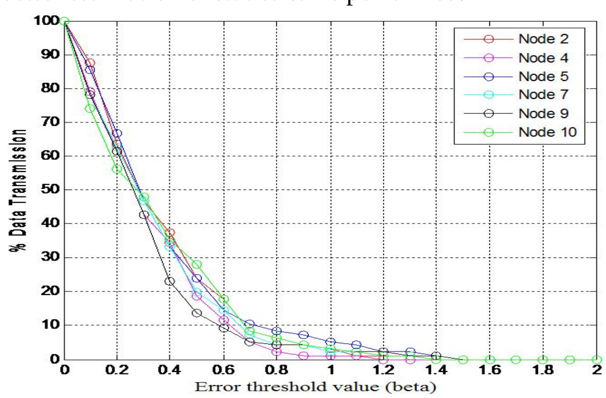

Figure 6. Percentage of data transmission versus error threshold value (β) .

Tracing malicious nodes using STDP model: Finally using STDP model, we can find the number of malicious nodes in the network. Malicious node can sense inaccurate data readings .The signal variations of malicious nodes are much higher than normal nodes. In our network, assuming node Id's 5 and 9 are malicious and node Id's 2, 4, 7 and 10 are normal nodes. Fig.8 shows that variations of the weighted vector of node Id's 5 and 9 are much higher than the normal nodes. The weighted vector $(w_{mal\_i,k}^{new})$ of malicious nodes shows abnormal signal variations than the normal nodes. Such abnormal signal variation of the weighted vector of sensor node detected at the sink node is said to be malicious node.

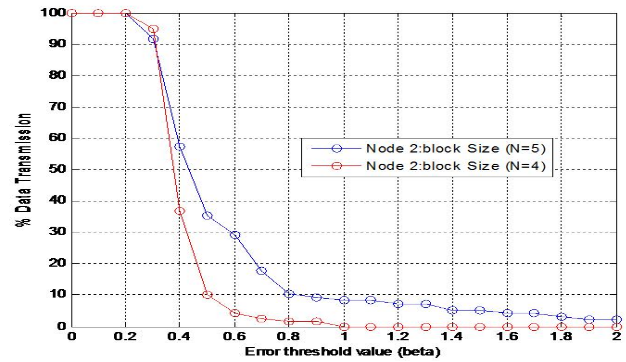

Figure 7. Percentage of data transmission versus error threshold value (β) for Block size $N = 4$ and $N = 5$ data streams

Thus we can easily trace the number of malicious nodes in the network by analyzing the signal variation of the weighted vector ($w_{i,k}^{new}$) of each node. The signal variation of weighted vector and the variance of each sensor node are summarizes in Table I. Thus from Table I , we conclude that the weighted vector of node Id's 5 and 9 shows abnormal statistical variations than the normal nodes. Finally node Id's 5 and 9 can be discarded from the network to get better data accuracy in the network when we don't have a priori knowledge of signal statistics of physical environment.

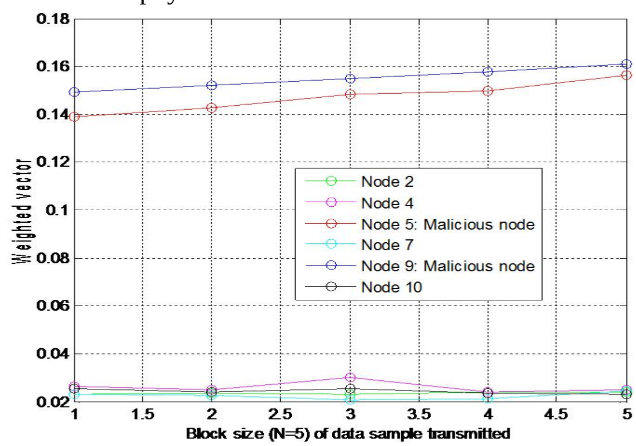

Figure 8. Weighted vector versus data sample block size (N=5) of each sensor node in STDP model to trace malicious nodes.

TABLE I: WEIGHTED VECTOR OF SENSOR NODES TO TRACE MALICIOUS NODES IN THE NETWORK.(*NOR:* NORMAL NODES AND *MAL:* MALICIOUS NODES)

Selecting Nodes	Node 2	Node 4	Node 5	Node 7	Node 9	Node 10
Weighted vector of sensor nodes to show the statistical behaviour using STDP model	0.0232	0.0264	0.1390	0.0229	0.1493	0.0253
	0.0236	0.0248	0.1429	0.0225	0.1520	0.0242
	0.0230	0.0302	0.1484	0.0207	0.1551	0.0253
	0.0240	0.0242	0.1500	0.0214	0.1575	0.0236
	0.0238	0.0250	0.1561	0.024	0.1609	0.0230
variance	1.3760e-07	4.6816e-06	3.4838e-05	1.6296e-06	1.6538e-05	8.3760e-07
Result	*Nor*	*Nor*	*Mal*	*Nor*	*Mal*	*Nor*

VI. Conclusions

In this paper, we presented *two* data accuracy models to sense accurate data from the physical environment. *First*, Estimated Data Accuracy (EDA) model which require a priori information of signal statistics of environment. EDA model senses more accurate data and performs better than other information accuracy models. *Second* Adaptive Data Accuracy (ADA) model to select an optimal sensor nodes in the network under adaptive approach. ADA model doesn't require any a priori knowledge of the signal statistics of the environment. Moreover we describe Spatio-Temporal Data Prediction (STDP) model which reduces the communication overhead for these optimal sensor nodes under data reduction strategies. Simulation results show that STDP can learn and track the internal variation of signal statistics of the environment. Finally we propose a mechanism using STDP model to trace the malicious nodes in the network if any, due to extreme physical environment e.g heavy rainfall. Extensive simulation results are performed to validate EDA, ADA and STDP models respectively under malicious network.

References

[1] I.F Akyuildz , W. Su ,Y. Sankarasubramanian ,E.Cayirci ,"A Survey on sensor Networks", IEEE Communications Magazine ,vol.40,pp.102-104 (Aug 2002).

[2] S.S Pradhan ,K.Ramchandran,"Distributed Source Coding: Symmetric rates and applications to sensor networks", proc. of the data compressions conference,pp.363-372 (2000).

[3] M.Gastpar, M.Vetterli ,"Source Channel Communication in Sensor Networks", Second International Workshop on Information Processing in Sensor Networks (2003).

[4] M.C Vuran,O.B Akan,I.F Akyildiz," Spatio-temporal correlation: theory and applications for WSNs",Computer networks, Elsevier,pp245-259 (2004).

[5] H.Li,S. Jiang , G. Wei ," Information accuracy aware jointly sensing nodes selection in wireless sensor networks", LNCS , Springer –Verlag Berlin , pp. 736-747 (2006).

[6] K.Cai,G. Wei, H. Li ," Information accuracy versus jointly sensing nodes in WSNs", IEEE Asia Pacific conference on cicuits and systems , pp.1050-1053 (2008).

[7] Jyotirmoy karjee , H.S jamadagni ," Optimal Node Selection using Estimated Data Accuracy Model in Wireless Sensor Networks",Proceedings of 3rd International conf. on Recent Trends in Information, Telecom. & Computing,LNEE,vol-150,pp-195-205,(2013).

[8] Jyotirmoy karjee , H.S Jamadagni ," Energy Aware Node Selection for Cluster-based Data accuracy Estimation in wireless sensor networks" International Journal of Advanced Networking and Applications ", vol-3 , Issue 5 , pp. 1311-1322 (2012).

[9] Jyotirmoy Karjee , H.S Jamadagni , " Data Accuracy Model for Distributed Clustering Algorithm based on Spatial Data Correlation in Wireless Sensor Networks",
GESJ:Computer Science and Telecommunications, No.2 (34),pp.13-27, (2012)

[10] Jyotirmoy Karjee , H.S Jamadagni ," Data accuracy Estimation for Spatially Correlated Data in Wireless Sensor networks under Distributed Clustering ", Journal of Networks , vol-6 , no.7, pp 1072-1083 (july 2011).

[11] Jyotirmoy Karjee , H.S Jamadagni ," Data accuracy estimation for cluster with spatially correlated data in wireless sensor networks",IEEE CSNT-2012,pp. 427-434,(2012).

[12] Jyotirmoy karjee , Sudipto Banerjee , " Tracing the Abnormal Behavior of Malicious Nodes in MANET ", Fourth International Conference on Wireless Communications , Networking and Mobile Computing , Dalian , China, pp. 1-7(2008).

[13] T.J. Goblick , " Theoretical Limitations on the transmission of data from analog sources ," IEEE Transaction Theory , IT-11 (4) pp.558-567 (1965).

[14] V.Poor ," An Introduction to signal Detection and Estimation ", Second Edition , Springer , Berlin(1994).

[15] A.H Syed, Adaptive Filters, John Wiley & Sons, NJ , (2008).

[16] J.O Berger , V.de Oliviera and B. Sanso , " Objective Bayesian Analysis of Spatially Correlated Data ", J.Am. Statist. Assoc. Vol-96, pp.1361-1374(2001).

[17] De Olivera V, Kedan B and Short D.A , " Bayesian prediction of transformed Gaussian Random Fields ", Journal of American Statistical Association,vol-92,pp.1422-1433(1997)

[18] M.A. Perillo ,W.B. Heinzelman ," Optimal sensor management under energy and reliablity constraints ", Proc. of IEEE wireless communications and networking, pp. 1621-1626(2003).

[19] F.Ordonez and B.Krishnamachari, "Optimal information extraction in energy limited wireless sensor network," IEEE Journal of selected area in communications, vol. 22, no.6 , pp. 1121-1129 (August 2004).

[20] T.Panigrahi,P.M. Pradhan,G.Panda and B.Mulgrew ," Block Least Square Algorithm over Distributed Wireless Sensor Network", Journal of Computer Networks and Communications ", vol-2012, Article ID : 601287(2012).

[21] C.C. Lopes and Ali H.Syed ,"Distributed Processing over Adaptive Networks " , Ninth International Symposim on Signal Prccessing and its Applications (ISSPA),pp.1-3 , 2007

[22] C.C Lopes and Ali H Syed ," Incremental Adaptive Strategies Over Distributed Networks ", IEEE Transaction on Signal Processing , vol.55 ,pp.4064-4077,(Augest 2007).

[23] B.Q Ali , N. Pissinou , K. Makki , " Estimated Replication of Data in Wireless Sensor Networks ", Third International Confernce on Communication System software and Middleware Workshop. pp. 107-110, 2008.

[24] B.Q Ali , N. Pissinou , K. Makki , " Approximate Replication of Data Using Adaptive Filters in Wireless Sensor Networks ", Third International Symposium on Wireless Pervasive Computing (ISWPC), pp. 365-369,2008.

[25] S.Santini ,K. Romer " An Adaptive Strategy for Quality – based Data Reduction in Wireless Sensor Networks ",In proceedings of 3rd International Conference on Networked Sensing Systems (INSS'06) , USA , June 2006.

[26] A. Deshpande , C. Guestrin , S.R Madden, J.M Hellerstein and W.Hong , " Model Driven Data Acquisition in sensor networks ", In proccedings of 30th VLDB confernce,Toronto, Canada, 2004.

[27] B.R Stojkoska , D. Solev,D.Davcev, "Variable step size LMS Algorithm for Data Prediction in wireless sensor networks",Sensor & Transducers Journal ,vol. 14-2 ,Special Issue,pp. 111-124,Feb 2012 .

[28] Weather history for Jaisalmer, India: http://www.wunder ground.com

[29] Mawsynram Climate and rainfall : http://en.wikipedia.org / wiki/Mawsynram